Interleaving Schemes on Circulant Graphs

Aleksandrs Slivkins
Department of Computer Science
Cornell University
Ithaca, NY 14853
Email: slivkins@cs.cornell.edu

Jehoshua Bruck
Department of Electrical Engineering
California Institute of Technology
Mail Code 136-93, Pasadena, CA 91125
Email: bruck@paradise.caltech.edu

Abstract

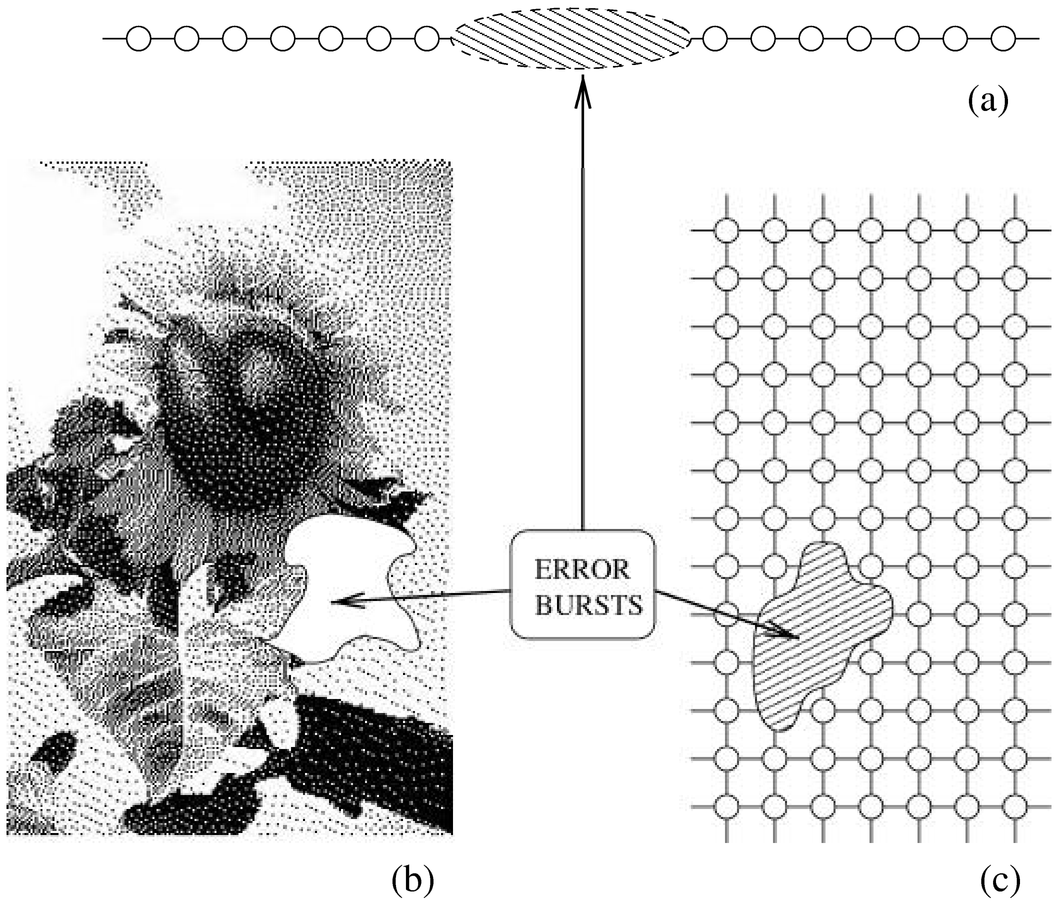
Interleaving schemes are used for error-correcting on a noisy channel. We consider interleaving schemes on infinite circulant graphs with two offsets 1 and d, with a goal to minimize the interleaving degree. Our constructions are minimal covers of the graph by copies of some subgraph S that can be labeled by a single label. We focus on minimizing the *index* of S – an inverse of its density rounded up. We establish lower bounds and prove that our constructions are optimal or almost optimal, both for the index of S and for the interleaving degree. We identify related combinatorial questions and advance conjectures.

1 Introduction

Error-correcting codes work best when the errors are scattered. However, in real life errors on noisy channels often are bursty. So *interleaving* is used. The idea is to assign data points to a number of separate codes, so that the points assigned to the same code are not likely to be hit by one and the same error burst. T hen from eachode's perspective there would be no or few bursts. The goal is to minimize the number of distinct codes.

For a simple example, suppose we transmit bits of information one by one and we'd like to use parity bits for error-correcting. Furthermore, suppose we know that error bursts are quite rare, but a single error burst can damage up to three consecutive bits. So we split the bits into three groups as 123123123123.... For each group, we include one parity bit per, say, N consecutive bits transmitted. Then a single error burst can damage at most one bit in each group. The transmission overhead is proportional to the number of groups (distinct codes).

The way we interleave the codes largely depends on the topology of a noisy channel. Usually noisy channels are one-dimensional, like a wireless data link, time being the only dimension. In this case error bursts are just segments of the line (Fig. 1a). If we draw a picture on a piece of paper and then scan it, we can view the piece of paper as a noisy channel that transmits data from us to the scanner (Fig. 1b). Here error bursts are two-dimensional (Fig. 1c). A holographic data storage system can be viewed as a three-dimensional noisy channel (see p.1 of [1]).



(a) One-dimensional noisy channel.
(b) A blot on a picture.
(c) Model of a blot: error burst in two dimensions.

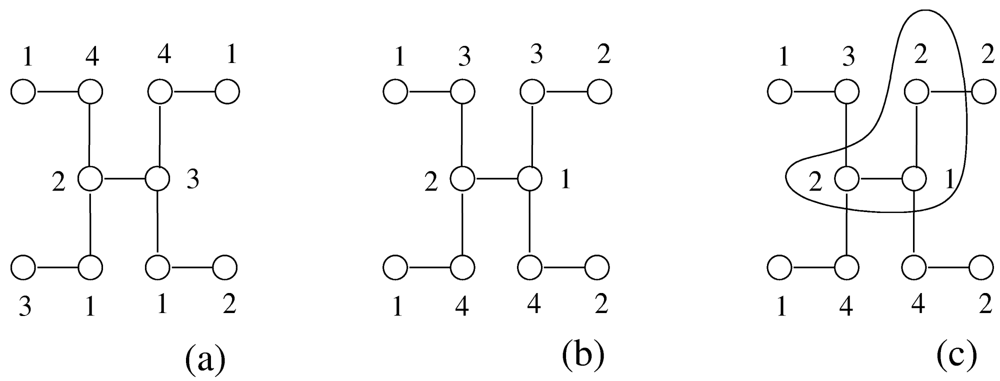
Figure 1: Error bursts in one and two dimensions

Formally the topology of a noisy channel is given by a graph G on data points transmitted by the channel. For example, the topology of an n-dimensional noisy channel is an n-dimensional mesh. The idea is to model error bursts as connected subgraphs of G, i.e. to make sure that two data points are likely to be hit by the same error burst iff they are close to each other in G. In particular, we usually connect each data point with data points transmitted immediately before and after it. Note that this is but a mathematical model, i.e. an approximation. As such, it can be good or bad, but not 'right' or 'wrong'.

Methods of interleaving are typically designed for specific topologies and specific types of error-bursts, with specific performance guarantees: *this code with this many codewords corrects for any burst of this specific kind.* A numerical measure on error-bursts is usually defined. Then the number of codewords is expressed as an increasing function of this measure. For example, we could consider two-dimensional meshes, restrict the shapes of error bursts to rectangles $t_1 \times t_2$, and use $t_1 + t_2$ for the measure. Then our performance guarantee would have the following form: our code needs $f(t_1 + t_2)$ codewords to correct for any rectangular burst $t'_1 \times t'_2$ s.t. $t'_1 + t'_2 \leq t_1 + t_2$. Different shapes and measures of bursts have been studied, mostly on one- and two- dimensional meshes (see [1] for references).

By the *size* of an error burst we mean the number of bits corrupted by it. The present paper takes after [1] in that it tries to minimize the number of codewords required to correct for *any* error burst of a given size t. In other words, we need to make sure no error burst of size t or less contains two data points assigned to the same code.

In terms of graphs, error bursts are connected subgraphs, and assigning data points to codes is just a labeling. So we have a labeling problem: given a graph G and an integer t, construct a labeling of G so that no connected subgraph of size t contains two vertices labeled the same (equivalently, the distance between any two vertices labeled the same is at least t). Such a labeling is called a *t-interleaving scheme*, where t is an *interleaving parameter* (Fig. 2). The goal is to minimize *interleaving degree*, the number of distinct labels (codes) used. Note that for $t = 2$ it is just the graph-coloring problem.



(a)&(b) 3-interleaving schemes with interleaving degree 4.
(c) Not a 3-interleaving scheme.

Figure 2: Examples of interleaving schemes

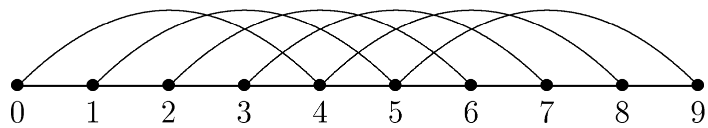
The history of the work on interleaving schemes is rather brief. The original paper [1] introduced interleaving schemes and analyzed them on two- and three-dimensional arrays. In [2], interleaving schemes were generalized to interleaving schemes with repetitions, where in every connected cluster of size t any label is repeated at most r times. Asymptotically optimal constructions on 2-dim arrays were presented for the case r=2. Paper [4] considered the case $r > 2$.

Present scope

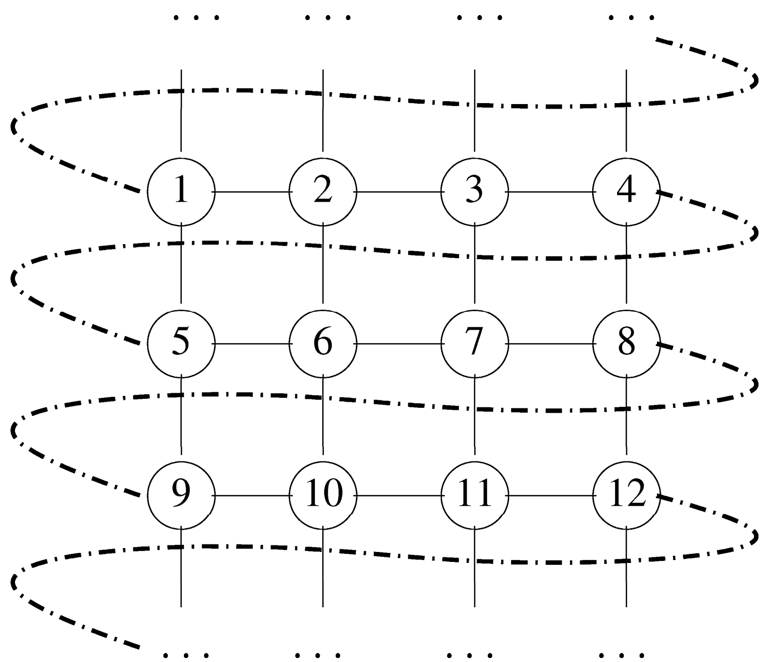
The problem of minimizing the interleaving degree is two-fold: constructions and lower bounds; we address both. For us the problem is more combinatorial than practical; we are especially interested whether and when our constructions and bounds are *exactly* optimal.

In this paper we restrict ourselves to infinite circulant graphs G_d with two offsets $\{1, d\}$. G_d is a graph on $\mathbb{Z}$ s.t. vertices i,j are connected iff $|i - j|$ is 1 or d (Fig. 3a). We talk interchangeably about subgraphs of G_d and subsets of $\mathbb{Z}$.

The motivation for choosing G_d is that it is topologically similar to a two-dimensional rectangular mesh (Fig. 3b), the topic of the previous papers on interleaving schemes. Note that G_d is embeddable on a cylinder: embed the integer line as a spiral.



(a) G_4: the usual view. The graph extends infinitely to both sides.



(b) G_4 as a two-dim mesh with a few extra edges.

Figure 3: G_4, the infinite circulant graph with two offsets $\{1, 4\}$.

All points of a set S can be labeled by a single label in an interleaving scheme iff the distance between any two of them is at least t. Call such sets t-sparse. An interleaving scheme with n labels $L_1 \dots L_n$ can be viewed as partition of the graph into t-sparse sets $C_1 \dots C_n$ s.t. all vertices of C_i are labeled by L_i. Actually, we just need C_i's to cover the graph: if C_i intersects C_j, points in the intersection can be labeled L_i or L_j, arbitrarily.

Our approach is to cover $\mathbb{Z}$ with copies of a periodic t-sparse set C. For periodic sets density is defined as the number of points within the period over the length of the period. The *index* of a set is defined as the inverse of density rounded up. Note that the index of a set gives a lower bound on the number of copies of this set needed to cover $\mathbb{Z}$. We separate the problem of minimizing the interleaving degree into two problems:

- Find a t-sparse set C with a minimal index.
- Cover $\mathbb{Z}$ with the minimal number of copies of C.

Most of our progress is on minimizing the index of a t-sparse set, which is itself an interesting combinatorial problem.

This approach is not guaranteed to yield the smallest interleaving degree. In particular, it might be possible to cover $\mathbb{Z}$ with fewer copies of a t-sparse set with a larger index. Our constructions are optimal sometimes and good approximations otherwise.

Our results

There are several cases which require separate constructions and lower bounds.

For $t \geq d-1$ there is a simple unique optimal interleaving scheme and t-sparse set.

For each choice of (d,t) s.t. $\lceil d/2 \rceil < t \leq d-2$ we present a family of optimal t-sparse sets and extend one of them to an interleaving scheme that is optimal in about half of the cases and a $(1+\frac{4}{t})$-approximation otherwise.

For $d \geq 2t$ our "practical" results are:

- a *sphere-packing* lower bound (SLB) based on packing of $\mathbb{Z}$ by copies of a t-sphere[1].
- two optimal constructions for sparse but infinite subsets of pairs (d,t).
- a construction for all (d,t) that is a $(1+\frac{t}{d}+\frac{1}{t})$-approximation of SLB.
- for even t s.t. $d > t^3$ we construct t-sparse sets with index just one above SLB.

Aside from that,

- for odd t we determine precisely when SLB for t-sparse sets is exact.
- We investigate when $w\mathbb{Z}$, $w \in \mathbb{N}$ is an optimal t-sparse set.
- We formalize and study the "greedy" approach for constructing t-sparse sets.

Further research

The two natural generalizations of interleaving schemes on G_d are interleaving schemes with repetitions and interleaving schemes on circulant graphs with more than two offsets (Fig. 4). Aside from that, in the end of each section we identify open questions.

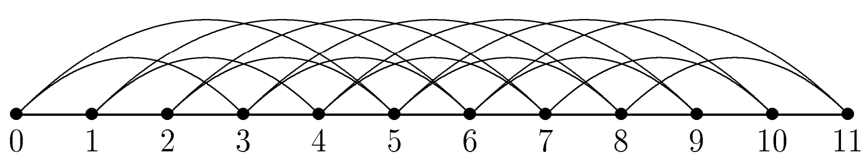


Figure 4: Infinite circulant graph with three offsets $\{1,3,5\}$.

Organization of this paper

The paper is divided into 5 sections of approximately equal size:

Section I Introduction.

Section II A sphere-packing lower bound (SLB) on the index of a t-sparse set.

Section III Our constructions for the case $t \leq \lceil d/2 \rceil$.

Section IV When is SLB exact? When is $w\mathbb{Z}$, $w \in \mathbb{N}$ is an optimal t-sparse set?

[1] t-sphere is a version of a "usual" sphere to be defined in Section II.

Section V Our constructions and lower bounds for the case $t > \lceil d/2 \rceil$.

The main flow of the paper is about lower bounds and constructions for periodic t-sparse sets and interleaving schemes. In the end of each section we present conjectures and open questions and discuss results that would otherwise interrupt the main flow.

2 Sphere-Packing Lower Bound

We look for lower bounds on the index of a t-sparse set. The idea is to express t-sparseness via packing of isomorphic subgraphs. We define a version of a sphere on G_d [2] called a t-sphere. It turns out that if a set S is t-sparse then the t-spheres centered in the points of S are pairwise disjoint (form a packing),[3] so the index of S is equal or greater than the size of a t-sphere. We call this lower bound a *sphere-packing* lower bound (SLB). Our constructions either reach SLB or come close. In Section IV we investigate whether and when SLB is exact.

We only consider the case $t \le \lceil d/2 \rceil$. In section V we'll show that for other cases there are better lower bounds based on different ideas.

We reserve d for the larger offset in G_d, and t for the interleaving parameter. We write S_t for a t-sphere. For given d, t we denote the minimal interleaving degree of a periodic interleaving scheme by $\text{DEG}_{\min}$, and the minimal index of a periodic t-sparse set by $\text{IND}_{\min}$. Now we can state the main result of this section.

Theorem 2.1 (Sphere-packing lower bound) $\text{DEG}_{min} \ge \text{IND}_{min} \ge |S_t| = \lceil t^2/2 \rceil$. *Moreover,* $\text{IND}_{min} = |S_t|$ *iff* $\text{DEG}_{min} = |S_t|$.

First we define t-spheres abstractly and prove that their size bounds $\text{IND}_{\min}$. Then we investigate what they actually look like and compute their size.

Define the distance between points u, v as the number of edges in a shortest uv-path; denote it by $dist(u, v)$. Let $dist(v) = dist(0, v)$. Define the distance between a subgraph S and a point v as the minimal distance between v and the elements of S. Denote it by $dist(S, v)$. For an integer τ define the τ-span of a set S as the set of points at distance less than t from S. Denote it by $\text{SPAN}_\tau(S)$.

Definition 2.1 *Define three kinds of t-spheres centered at $p \in G_d$ as*

$$\begin{aligned} S_{2\tau-1}(p) &= \text{SPAN}_\tau\{p\} \\ S^S_{2\tau}(p) &= \text{SPAN}_\tau\{p, p+1\} \\ S^L_{2\tau}(p) &= \text{SPAN}_\tau\{p, p+d\} \end{aligned}$$

For clarity we'll call $S^S_{2\tau}$ a *short* t-sphere, and $S^L_{2\tau}$ a (*long*) t-sphere. Some examples of t-spheres are in Fig. 2. By the end of this section it will become clear why t-spheres look the way they look.

Lemma 2.1 *If $dist(p, q) \ge t$ then*

$$\begin{cases} S_t(p) \not\pitchfork S_t(q), & t \text{ odd,} \\ S^S_t(p) \not\pitchfork S^S_t(q) \text{ and } S^L_t(p) \not\pitchfork S^L_t(q), & t \text{ even} \end{cases}$$

Proof: Suppose $t = 2\tau - 1$ and $S_t(u_1)$ and $S_t(u_2)$ intersect at w. Then $dist(u_i, w) \le \tau - 1$, so by triangle inequality $dist(u_1, u_2) < t$.

Now let $t = 2\tau - 2$ and suppose $S^S_t(u_1)$ and $S^S_t(u_2)$ intersect at w. Then for each i either $dist(u_i, w) \le \tau - 1$ or $dist(u_i + 1, w) \le \tau - 1$. Therefore, by triangle inequality $dist(u_1, u_2) < t$

[2] As before, G_d stands for a circulant graph with two offsets $\{1, d\}$

[3] The inverse holds for odd t only. For even t see Lemma 4.5.

(a) Let S be a t-sphere centered at $p \in G_d$ (i.e. S is S_t or S_t^S or S_t^L). We represent S by a string where consecutive characters correspond to consecutive numbers as follows.

♠	for the points $p + kd \in S$, $k \in \mathbb{Z}$;
◇	for the points $p + 1 + kd \in S$, $k \in \mathbb{Z}$ (only for $S_t^S(p)$);
•	for other points of S;
x	for points not in S.
◯	to label the centers.

(b) The three kinds of t-spheres for $d = 7$: $\tau = 3$ (above) and $\tau = 4$ (below).

$S_{2\tau}^S$	♠◇xxxx•♠◇•xx••♠◇••xx•♠◇•xxxx♠◇
$S_{2\tau-1}$	♠xxxxx•♠•xxx••♠••xxx•♠•xxxxx♠
$S_{2\tau}^L$	♠xxxxx•♠•xxx••♠••xx••♠••xxx•♠•xxxxx♠
$S_{2\tau}^S$	♠◇xxxx•♠◇•xx••♠◇•••••♠◇•••••♠◇••xx•♠◇•xxxx♠◇
$S_{2\tau-1}$	♠xxxxx•♠•xxx••♠••x•••♠•••x••♠••xxx•♠•xxxxx♠
$S_{2\tau}^L$	♠xxxxx•♠•xxx••♠••x•••♠••••••♠•••x••♠••xxx•♠•xxxxx♠

Figure 5: t-spheres

unless $dist(u_1, w) = dist(u_2, w) = \tau$. In the latter case, however, $dist(u_1 + 1, w) = dist(u_2 + 1, w) = \tau - 1$, so there exists a path $u_1 + 1 \to u_2 + 1$ with $\leq t$ vertices. Shifting it by -1 produces a $u_1 u_2$ path of the same length, so that, again, $dist(u_1, u_2) < t$. Similar proof works for long t-spheres, too. $\square$

Corollary 2.1 *$\text{IND}_{min} \geq |S_t|$, t is odd. $\text{IND}_{min} \geq \max(|S_t^S|, |S_t^L|)$, t is even.*
Proof: Let S be a t-sparse set, t is odd. Then $\mathcal{F} = \{S_t(p) : p \in S\}$ is a family of disjoint sets. Let ρ be the density of S. Then the density of the union of all sets in $\mathcal{F}$ is $\rho|S_t| \leq 1$. Thus, $\rho \leq |S_t|^{-1}$, so the index of S is at least $|S_t|$.

Similar arguments work for even t. $\square$

Definition 2.2 *If a t-sparse set or an interleaving scheme reaches* SLB*, call it* s-optimal.

Lemma 2.2 *S-optimal t-sparse set gives rise to an s-optimal t-interleaving scheme.*
Proof: Let S be an s-optimal t-sparse set Let W be the union of t-spheres centered in S. These t-spheres are disjoint, so the density of W is 1. Since S is periodic, W is periodic, too. Thus, $W = \mathbb{Z}$.

Partition W as follows: let $W_i = \{f_i(p) : p \in S\}$, where $f_i(p)$ is the i^{th} vertex of $S_t(p)$ from the left. Then the sets W_i are translates of S, hence t-sparse. Label all points of W_i with i to get an s-optimal interleaving scheme. $\square$

As a result of this abstract discussion of t-spheres we have almost proved Thm. 2.1, except at this point we have no idea what these t-spheres are like, and what is their size. We'll have to take a closer look at G_d.

Some formulas will look a bit different for odd d and for even d. To avoid writing similar things twice we introduce $\delta = \lceil d/2 \rceil$.

Definition 2.3 *Define a canonical representation* $\text{CAN}(v)$ *of* $v > 0$ *as a pair* (x, y) *s.t.* $v = xd + y$, $x \geq 0$ *and* $-\delta < y \leq \delta$.

By writing $\text{CAN}(v)$ we imply $v > 0$. Note that $\text{CAN}(v)$ is uniquely determined by v.

Lemma 2.3 *Suppose* $\text{CAN}(v_1) = (x_1, y_1)$, $\text{CAN}(v_2) = (x_2, y_2)$, *and* $x_1 > x_2$. *Then* $v_1 > v_2$.

Lemma 2.4 *Let* $\text{CAN}(v) = (x, y)$. *Then* $dist(v) = x + |y|$.

Proof: Define a canonical path P_l^s as a path in G_d that starts with 0 and consists of $|l|$ large jumps, followed by $|s|$ small jumps, in the direction determined by l, s, respectively (see Fig. 6). Since two jumps of the same size but in opposite directions can be removed, any shortest path from 0 is canonical.

Say P_l^s is a shortest path from 0 to some $v \in G_d$. If $l < 0$ then P_0^v is shorter. If $s > \delta$ then P_{l+1}^{s-d} is shorter. If $s \leq -\delta$ then P_{l-1}^{d-s} is shorter. Therefore $\text{CAN}(v) = (l, s)$. □

Corollary 2.2 *If* $v > 0$ *then* $dist(v + d) = dist(v) + 1$.

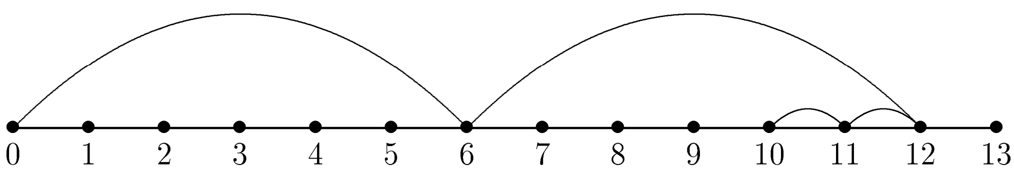


Figure 6: P_2^{-2} is a shortest path from 0 to 10 in G_6.

Definition 2.4 *Define the positive and negative* τ*-spans (Fig. 7)*

$$\begin{aligned} \text{SPAN}_\tau^+ &= \{p \geq 0 \mid dist(p) < \tau\} \\ \text{SPAN}_\tau^- &= \{p \leq 0 \mid dist(p) < \tau\} \end{aligned}$$

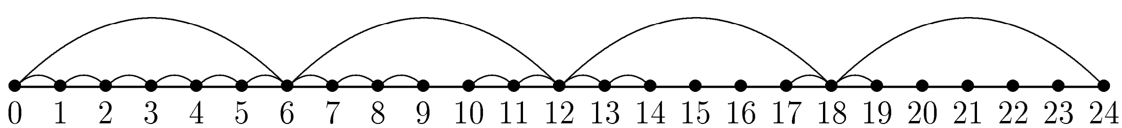


Figure 7: SPAN_5^+ on G_6.

Definition 2.5 *Let* v_{max}^τ *be the smallest positive number in* SPAN_τ^+.

Obviously, $\text{SPAN}^-_\tau = -\text{SPAN}^+_\tau$ and $S_{2\tau-1} = \text{SPAN}^+_\tau \cup \text{SPAN}^-_\tau$

We will represent a SPAN^+_τ more compactly with a string where characters are in 1-1 correspondence with integers in $[0; v^\tau_{\max}]$:

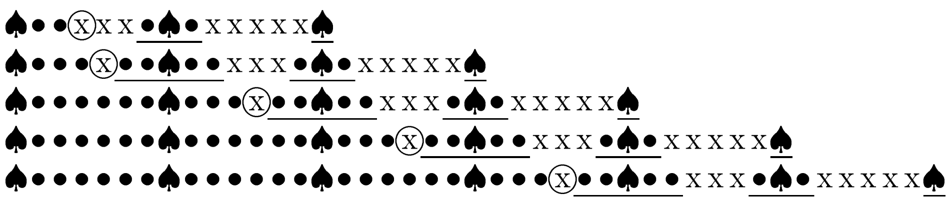
♠ for the multiples of d,

• for other element of SPAN^+_τ,

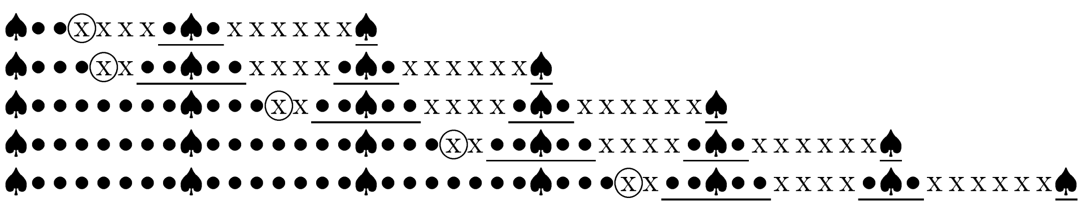
x for points not in SPAN^+_τ,

◯ to label some of the vertices.

Fig. 8 shows how SPAN^+_τ grows with τ for a fixed d.



(a) SPAN^+_τ for $d = 7$, $\tau = 3$-7. $v^\tau_{\min}(\tau)$ is labeled.



(b) SPAN^+_τ for $d = 8$ and $\tau = 3$-7. $v^\tau_{\min}$ is labeled.

Figure 8: For a fixed d SPAN grows with τ. Stations are underlined.

Lemma 2.5 $v^\tau_{max} = d(\tau - 1)$.

Proof: Let $\text{CAN}(v^\tau_{\max}) = (x, y)$. If $y \neq 0$ then, for $v = d(x + |y|)$, $dist(v) = dist(v^\tau_{\max})$ but $v > v^\tau_{\max}$, contradiction. So wlog $v^\tau_{\max} = xd$, with $dist(v^\tau_{\max}) = x$. □

Definition 2.6 *Let v^τ_{min} be the smallest positive number not in SPAN^+_τ.*

Lemma 2.6 $v^\tau_{min} = \begin{cases} \tau, & t \leq \delta \\ (\tau - \delta)d + \delta, & t \geq \delta \end{cases}$

Proof: Let $\text{CAN}(v^\tau_{\min}) = (x, y)$. If $dist(v^\tau_{\min}) > \tau$ then $dist(v^\tau_{\min} - d) \geq t$, contradiction. Thus, $x + |y| = \tau$. By Lemma 2.3, $x \geq 0$ must be as small as possible, so $|y|$ must be as large as possible. Lemma follows because $|y| \leq \delta$ and $y > -\delta$. □

Note the structure of SPAN^+_τ between $v^\tau_{\min}$ and $v^\tau_{\max}$. Elements of SPAN^+_τ are clustered around the multiples of d. We will call these clusters *stations*: imagine a train going in the positive direction with stops at $0, d, 2d, \ldots$, each stop serving a certain area. By lemma 2.4, the station $\mathcal{T}_k$ centered at kd is the integer interval $kd + [-x, x]$, where $x = \tau + k + 1$.

Let's return to t-spheres. Recall that SPAN_τ^+ is the right half of $S_{2\tau-1}$. Define stations $\mathcal{T}_k$ for $k \leq 0$ as $\mathcal{T}_k = -\mathcal{T}_{-k}$. Some of the stations might intersect. It turns out the sphere-packing lower bound is of interest only when all stations are disjoint, i.e. when $v_{\min}^\tau = \tau$ where $\tau = \lceil t/2 \rceil$ (equivalently, when $t \leq \delta$). We'll compute the size of a t-sphere for this case only.

Lemma 2.7 *For odd t, $|S_t| = \lceil t^2/2 \rceil$. For even t, $|S_t^S| = |S_t^L| = \lceil t^2/2 \rceil$.*
Proof: For odd t, S_t is a disjoint union of stations. Since $|\mathcal{T}_k| = 2(\tau - k) - 1$,

$$|S_t| = \sum_{k=-(\tau-1)}^{\tau-1} |\mathcal{T}_k| = (t^2+1)/2$$

For even $t = 2\tau$,

$$\begin{cases} S_t^S &= S_{t-1}(0) \cup S_{t-1}(1) \\ S_t^L &= S_{t-1}(0) \cup S_{t-1}(d) \end{cases}$$

Letting

$$\begin{cases} S &= S_{t-1}(1) \setminus S_{t-1}(0) \\ S' &= S_{t-1}(d) \setminus S_{t-1}(0) \end{cases}$$

we get, letting $x = \tau - 1$, (see Fig. 2)

$$\begin{aligned} S &= \{1 + \max(\mathcal{T}_k) \mid k \in [-x; x]\} \\ S' &= \{\min(\mathcal{T}_k) - 1, \max(\mathcal{T}_k) + 1 \mid k \in [1; x]\} \cup \{\tau d\} \\ |S| &= |S'| = t - 1 \\ |S_t^S| &= |S_t^L| = |S_{t-1}| + |S| = t^2/2 \end{aligned}$$

as required. □

Discussion and open questions

Definition 2.7 *Sub-graph $S \subset G_d$ is t-tight iff $\forall\ u, v \in S\ \mathit{dist}(u, v) < t$.*

Since no two points of a t-tight set can be labeled the same in an interleaving scheme, the size of a t-tight set gives a lower bound on the interleaving degree (but not on $\text{IND}_{\min}$).

Lemma 2.8 *t-spheres are are t-tight.*
Proof: Set $\{p\}$ is 1-tight; sets $\{p, p+1\}$ and $\{p, p+d\}$ are 2-tight. By Def. 2.1 it suffices to prove that if a set S is x-tight then $\text{SPAN}_y(S)$ is $(x + 2y - 2)$-tight. Indeed, for any $p, q \in \text{SPAN}_y(S)$, there are $p', q' \in S$ s.t. $\mathit{dist}(p, p')$ and $\mathit{dist}(q, q')$ are at most $y - 1$. Since $\mathit{dist}(p', q') < x$, by triangle inequality $\mathit{dist}(p, q) < x + 2y - 2$. □

Therefore, SLB holds for non-periodic interleaving schemes, too.

We'll see in Section III that for all but a sparse family of cases there is a $O(t)$ gap between SLB and our construction. Therefore, it would be nice to produce a t-tight set larger than a t-sphere. However, it is an open question whether such set exists.

3 Greedy Approach and Two-Offset Construction

In the previous section we derived a sphere-packing lower bound (SLB). Now we'll use the *greedy* approach to construct t-sparse sets that reach or almost reach SLB, and extend them efficiently to interleaving schemes.

All things being equal, we prefer t-sparse sets with a simple structure, since they are "nicer", easier to implement and to reason about. The latter is important because extending a given t-sparse set to an interleaving scheme efficiently can be a challenging problem.

Algorithm 3.1 *Greedy algorithm.*

Start with a set $S = \{0\}$. For each $j = 1, 2, 3, \ldots$, insert j into S if $S \cup \{j\}$ is t-sparse. Let $\mathcal{G}^A$ be the set produced by the greedy algorithm.

Definition 3.1 *A set $S \subset \mathbb{N}$ is periodic, with a period $x \in \mathbb{N}$, if for all $n \in \mathbb{N}$*

$$n \in S \iff n + x \in S \tag{1}$$

S is right-periodic, with a period $x \in \mathbb{N}$, starting with $y \in \mathbb{N}$, if (1) holds for all $n \geq y$.

Lemma 3.1 *$\mathcal{G}^A$ is right-periodic.*

Proof: Let $T = T_k$ be the set containing the first k elements of $\mathcal{G}^A$. Let j be the next element of $\mathcal{G}^A$. Then j depends only on the elements of T between $M - dt$ and M, where $M = \max T$. Moreover, $j - M$ depends only on the *header* $H(T)$ of T

$$H(T) = \{i \in 0 \ldots dt \mid M - i \in T\} \tag{2}$$

Obviously, $H(T)$ determines the rest of S. Since there are only finitely many headers, there are integers $k < l$ s.t. $H(T_k) = H(T_l)$. Then $\mathcal{G}^A$ is periodic starting with $\max T_k$. $\square$

Definition 3.2 *The greedy construction $\mathcal{G}$ is obtained from $\mathcal{G}^A$ by replicating the (smallest) period of $\mathcal{G}^A$ in both directions.*

Obviously, the greedy construction $\mathcal{G}$ is t-sparse. The hope is that it is dense enough, because every $j \in \mathcal{G}$ is as close as possible to the smaller elements of $\mathcal{G}$. The problem is, of course, that maybe if we make some intervals in $\mathcal{G}$ larger, some subsequent intervals can be made shorter, thus increasing the overall density.

Define $q, r \in \mathbb{N}$ by $d = (q+1)t + r$, $-1 \leq r \leq t-2$. The first few numbers produced by the greedy algorithm are $\mathcal{G}_0 = \{0, t, 2t, \ \ldots \ , qt\}$. The next number is

$$w = \min\{w > qt \mid w \notin \textsc{span}_t(\mathcal{G}_0)\}$$

Obviously $w > (q+1)t$. We'll show later that $w \approx dt/2$.

Definition 3.3 *Define a two-offset set as a periodic set S with a period x s.t. $S \cup [0, x) = \mathcal{G}_0$. Say $\mathcal{G}^A$ is two-offset if $\mathcal{G}$ is two-offset with a period w.*

Sanity check: $\mathcal{G}^A$ is two-offset if it is right-periodic starting with 0, with a period w. Recall that t-sparse sets and interleaving schemes that reach SLB are called *s-optimal*.

Theorem 3.1 *$\mathcal{G}^A$ is two-offset iff $d \equiv 0, \pm 1 \pmod t$. If t is even and $d \equiv \pm 1 \pmod t$ then $\mathcal{G}$ is s-optimal.*

We'll prove Thm. 3.1 in Lemmas 3.2, 3.3 and 3.4. According to this theorem, if $d \equiv 0, \pm 1 \pmod t$, $\mathcal{G}^A$ and $\mathcal{G}$ are nice; their period is simply $\mathcal{G}_0$. Else, $\mathcal{G}^A$ is quite ugly. Computer search shows that the periods are rather long and lack apparent structure. We can prove that periodicity of $\mathcal{G}^A$ starts with some number $N > 0$, but we could not compute N as a function of (d, t).

Now we need to study the t-span of $\mathcal{G}_0$. For $0 \le j \le q$, let $v_i^j = id + jt$ and define the integer intervals $\mathcal{B}_i^j = v_i^j + (-x, x)$, where $x = t - i$. Let $v_i^{q+1} = v_{i+1}^0$, $\mathcal{B}_i^{q+1} = \mathcal{B}_{i+1}^0$ (see Fig. 9). The meaning of $\mathcal{B}_i^j$ is that it is the part of the t-span of v_0^j that lies in $[id - t, id + d + t]$.

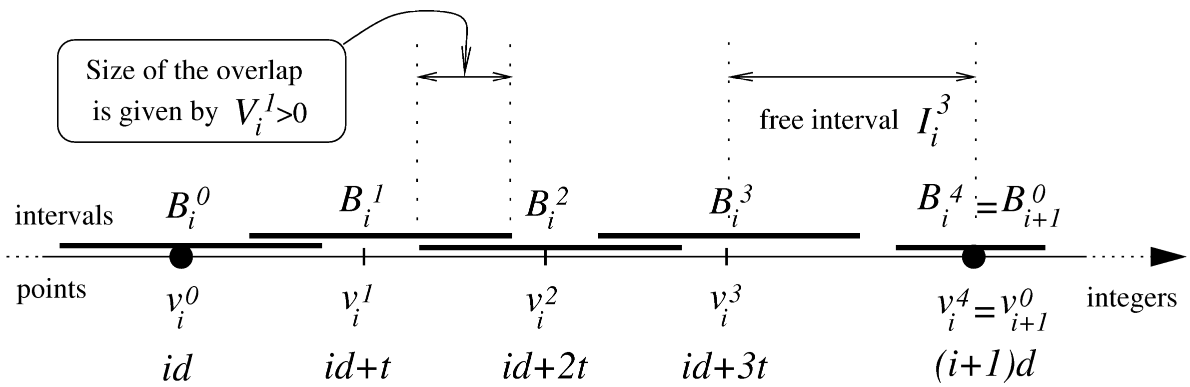


Figure 9: Notation: points v_i^j, intervals $\mathcal{B}_i^j$ and $\mathcal{I}_i^j$, ij-overlaps $\mathcal{V}_i^j$.

Say a point is $\mathcal{G}_0$-remote if its distance from $\mathcal{G}_0$ is at least t. Say a set is $\mathcal{G}_0$-remote if all its points are. Define the ij-overlap $\mathcal{V}_i^j$ as follows. If $\mathcal{B}_i^j$ and $\mathcal{B}_i^{j+1}$ overlap, let $\mathcal{V}_i^j$ be the size of the overlap. Else, let $-\mathcal{V}_i^j$ be the number of $\mathcal{G}_0$-remote points between $\mathcal{B}_i^j$ and $\mathcal{B}_i^{j+1}$. Then

$$\mathcal{V}_i^j = \begin{cases} t - 2i - 1, & 0 \le j < q \\ t - 2i - r - 2, & j = q \end{cases} \tag{3}$$

Partition the integer interval $id + [0, d)$ into integer intervals $\mathcal{I}_i^j = [v_i^j, v_i^{j+1})$, $0 \le j \le q$. Say $\mathcal{I}_i^j$ is free if it contains a $\mathcal{G}_0$-remote point. Sanity check: $\mathcal{I}_i^j$ is free iff $\mathcal{V}_i^j < 0$.

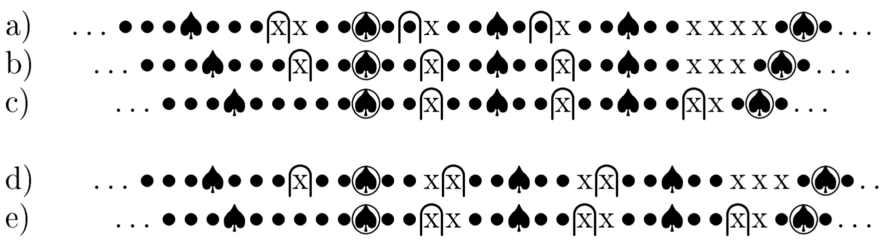
Call the interval $\mathcal{I}_i^j$ *standard* if $j < q$ and *non-standard* if $j = q$. Let $\tau = \lfloor t/2 \rfloor$, $\eta = \lfloor (t - r)/2 \rfloor$. Then by (3) $\mathcal{I}_\tau^0$ is the left-most standard free $\mathcal{I}_i^j$, and $\mathcal{I}_\eta^q$ is the left-most non-standard free $\mathcal{I}_i^j$.

Lemma 3.2 *If $r \ge 2$ then $\mathcal{G}^A$ is not two-offset.*

Proof: If $\eta < \tau - 1$, then $w \in \mathcal{I}_\eta^q$, so $w + t \in \mathcal{I}_{\eta+1}^0$ is not $\mathcal{G}_0$-remote since $\mathcal{I}_{\eta+1}^0$ is not free. Therefore, in this case $\mathcal{G}^A$ is not two-offset. Now, if $r \ge 3$ then $\eta < \tau - 1$; if $r = 2$ and t is odd, then $\eta = \tau - 2$. So it remains to consider the case $r = 2$ and t is even. Then $\eta = \tau - 1$, so we have to do some more arithmetic. By (3) there are exactly two $\mathcal{G}_0$-remote points in $\mathcal{I}_\eta^{q+1}$, namely w and $w + 1$, by definition of w (see Fig. 10a). The only one $\mathcal{G}_0$-remote point in $\mathcal{I}_\tau^0$ is $w + t + 1$. So again $w + t$ is not $\mathcal{G}_0$-remote. $\square$

In all examples G_0 consists of 3 elements. We represent the t-span of $\mathcal{G}_0$ as a string where consecutive characters correspond to consecutive integers as follows:

○	is for $d\tau$ and $d(\tau+1)$
♠	is for v_i^j's
•	for other elements of $\mathcal{B}_i^j$'s
x	for G_0-remote points
$\bigcap$	for w, $w+t$, $w+2t$.



a) $r=2$, t is even: $\mathcal{G}^A$ is not two-offset.
b-e) Typicals situations when $r \le 1$: $\mathcal{G}^A$ is two-offset.
b-c) t is even. b) $\eta = \tau - 1$; c) $\eta = \tau$.
d-e) t is odd. d) $\eta = \tau - 1$; e) $\eta = \tau$.

Figure 10: For the proofs of Lemmas. 3.2 and 3.3.

Observation 3.1 *All standard intervals $\mathcal{I}_\tau^j$ are free. Moreover, by (3) each such interval contains exactly one $\mathcal{G}_0$-remote point if t is even, and exactly two if t is odd. For all $j \le q$ the size of $\mathcal{B}_\tau^j$ is $t-1$ if t is even, and $t-2$ if t is odd.*

Lemma 3.3 *If $r \le 1$ then $\mathcal{G}^A$ is two-offset.*
Proof: Assume $r \le 1$. Then $\eta = \tau - 1$ or $\eta = \tau$, so by Obs. 3.1 $w + \mathcal{G}_0$ is $\mathcal{G}_0$-remote. See Fig. 10b-e for typical cases. The rest is a simple computation. Note that

$$w = \begin{cases} d\tau + \tau, & \eta = \tau \\ d\tau - \tau, & \eta = \tau - 1 \end{cases}$$

So for $j \le 3$ the set $jw + \mathcal{G}_0$ is $\mathcal{G}_0$-remote simply because it lies to the right of the t-span of $\mathcal{G}_0$. To complete the proof of Thm. 3.3, it remains to show $2w + \mathcal{G}_0$ is $\mathcal{G}_0$-remote, too.

The right-most point of the t-span of $\mathcal{G}_0$ is $p_1 = d(t-1) + qt$. The second right-most point is $p_2 = p_1 - t$. In particular, p_1 and p_2 lie to the left of dt. If $\eta = \tau$ then $2w > dt$. Suppose $\eta = \tau - 1$. If t is odd then again $2w > dt$. If t is even, then $2w = dt - t$. However, in this case $d(t-1) + qt = dt - t + 1$. So $2w$ lies between p_1 and p_2, and the rest of $2w + \mathcal{G}_0$ lies to the right of p_1. $\square$

The index of a two-offset set S with a period $w = d\tau \pm \tau$ is

$$\begin{aligned} \textsc{ind}(S) &= \left\lceil \frac{w}{q+1} \right\rceil = t\tau + \left\lceil \frac{(r \pm 1)\tau}{q+1} \right\rceil \qquad (4) \\ &= \begin{cases} |S_t| + \tau - 1 + \psi, & t \text{ is odd} \\ |S_t| + \psi, & t \text{ is even} \end{cases} \qquad (5) \end{aligned}$$

where ψ is the second summand in (4).

Lemma 3.4 *Suppose $r \leq 1$. Then $\mathcal{G}$ is s-optimal iff d is even and $r = \pm 1$, in which case it extends to an s-optimal interleaving scheme.*
Proof: Follows from (5) and Lemma 2.2.

As we have proved, the two-offset set with a period w is t-sparse only when $r \leq 1$. Does there exists a two-offset set with a large index that is t-sparse for *all* values of r? Indeed it does.

Definition 3.4 *Define the two-offset construction as a two-offset set with a period w_0, where $w_0 = d\tau + \tau$ is the $\mathcal{G}_0$-remote point (or the left-most of the two $\mathcal{G}_0$-remote points) in the interval $\mathcal{I}^0_\tau$.*

Sanity check: when $\mathcal{G}$ is two-offset and $w = d\tau + \tau$, it is exactly the same as the two-offset construction.

Lemma 3.5 *The two-offset construction is t-sparse.*
Proof: By Obs. 3.1 the set $w_0 + \mathcal{G}_0$ is $\mathcal{G}_0$-remote, As in the proof of Lemma 3.3, for $j \geq 2$ the set $jw_0 + \mathcal{G}_0$ lies to the right of the t-span of $\mathcal{G}_0$, hence is $\mathcal{G}_0$-remote.

By (4) the two-offset construction is a $(1 + \frac{t}{d} + \frac{1}{t})$-approximation of $\textsc{slb}$. If $d > t^3$, $\psi = 1$ in (5), so the index of the two-offset construction S is

$$\textsc{ind}(S) = \begin{cases} |S_t| + \tau, & t \text{ is odd} \\ |S_t| + 1, & t \text{ is even} \end{cases}$$

Now we will extend the two-offset construction to interleaving schemes.

Definition 3.5 *For a subset S of $\mathbb{Z}$, define the interleaving degree $\textsc{deg}(S)$ of S as the smallest number of copies of S required to cover $\mathbb{Z}$.*

Sanity check: if S is t-sparse then $\textsc{deg}(S)$ is the minimal interleaving degree of an interleaving scheme based on S.

Theorem 3.2 *Let S be a two-offset set with a period x. Let $g = \gcd(t, x)$. Then*

$$\textsc{deg}(S) = g \left\lceil \frac{x}{g(q+1)} \right\rceil = \begin{cases} \textsc{ind}(S), & \gcd(t,x) = 1 \\ \textsc{ind}(S) + g, & \textit{otherwise} \end{cases}$$

Proof: The second equality holds because $\textsc{ind}(S) = \lceil x/g \rceil$.

For a lower bound on $\textsc{deg}(S)$, instead of all $\mathbb{Z}$ let's just try to cover the integer interval $X = [0, 1)$. For $m \in \mathbb{Z}$, define the set

$$A_m = \{(m + jt) \bmod x : j \in \mathbb{Z}\}$$

From elementary number theory, the sets $A_0, A_1, ..., A_{g-1}$ form a disjoint partition of X, $|A_m| = x/g$. Now, each copy of S intersects with exactly one of the sets A_m, the size of intersection being $q+1$. Therefore, one needs at least $N = \left\lceil \frac{|A_m|}{q+1} \right\rceil$ copies to cover one of the sets A_m, and gN copies to cover all of them.

The covering of $\mathbb{Z}$ by gN copies of S is achieved by the sets

$$S_m^i = m + i(q+1)t + S$$

so that A_m is covered by $S_m^0, S_m^1, \ldots, S_m^N$. □

Corollary 3.1 *Let S be the two-offset construction.*

a) $\textsc{deg}(S)$ is a $(1 + \frac{t}{d} + \frac{3}{t})$-approximation of SLB *by (4).*

b) When $d > t^3$, $\textsc{deg}(S)$ is a $(1 + \frac{2}{t})$-approximation of SLB. *Specifically, by (5)*

$$\textsc{deg}(S) = \begin{cases} |S_t| + \tau + \gcd(d+1, t), & t \text{ odd} \\ |S_t| + \tau + 1, & t \text{ even, } d \text{ even} \\ |S_t| + t + 1, & t \text{ even, } d \text{ odd} \end{cases}$$

Discussion and open questions

1. Alg. 3.1 starts with an empty set S. This choice is quite arbitrary; instead, we can let the greedy algorithm start with any t-sparse set. Let $\mathcal{G}^A(S)$ be the (infinite) set produced by the greedy algorithm if it starts with S. Same way as when S was empty, one shows that $\mathcal{G}^A(S)$ is right-periodic and defines the greedy construction $\mathcal{G}(S)$. We do not know how dense the sets $\mathcal{G}(S)$ can be. It is an open question how the structure of the period of $\mathcal{G}(S)$ depends on S.

2. Given a t-sparse set S, the next number j_S produced by the greedy algorithm is the smallest number $j > \max S$ s.t. $S \cup \{j\}$ is t-sparse. As shown in the proof of Lemma. 3.1, j_S depends only on the header $H(S)$ of S, defined by (2). Therefore, one can view the greedy algorithm as a function $H(S) \to H(S \cup j_S)$. This function can be represented as a directed *transition graph* $G_\mathcal{G}$ on all possible headers. In particular, greedy construction $\mathcal{G}(S)$ corresponds to a directed cycle in $G_\mathcal{G}$ that contains, as a vertex, the header of $\mathcal{G}(S)$. Each connected component of $G_\mathcal{G}$ contains a directed cycle whose vertices are roots of disjoint trees directed towards the cycle. (Fig. 11). We'd like to know more about the structure of $G_\mathcal{G}$. For example, how many connected components can $G_\mathcal{G}$ have, for different values of (d, t)? how large are the cycles? what shape do the trees have? and so on.

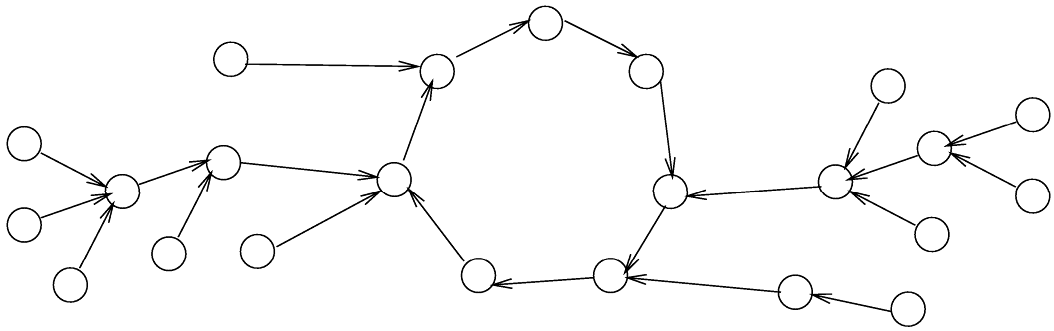


Figure 11: Typical connected component of a transition graph $G_{\mathcal{G}}$ of the greedy algorithm.

4 Even Constructions and S-Optimality

In this section we investigate when exactly s-optimal sets exist, and when the set $w\mathbb{Z}$, $w \in \mathbb{N}$ is s-optimal. It turns out that these questions are related.

The 'simplest' integer subsets are of the form $w\mathbb{Z}$, $w \in \mathbb{Z}$. If such subset is t-sparse, it gives rise to an interleaving scheme with interleaving degree w. Define the even construction as the set $|S_t|\mathbb{Z}$. Note that among all sets $w\mathbb{Z}$ only the even construction can be s-optimal.

Now we are ready to state the main results of this section:

- For odd t, s-optimal constructions exist iff $d \equiv \pm t \pmod{|S_t|}$, in which case the even construction is s-optimal.
- For even t, the even construction is s-optimal only if $d \equiv \pm 1 \pmod{t}$. Experimental data suggests some conjectures for when exactly the even construction is s-optimal.

The t-spheres centered in an s-optimal set partition $\mathbb{Z}$. Our approach for finding necessary conditions for s-optimality is to look at the partitions of $\mathbb{Z}$ by t-spheres.

In this section, let S be a (general) s-optimal set; let p be a (general) element of S. The notation for picturing t-spheres is summarized in Fig. 2a.

Case of odd $t = 2x - 1$

Fig. 12 shows that we can partition a t-sphere into *stations* (clusters around the points of the form $p + kd$, $k \in \mathbb{Z}$), which can be grouped into a *left branch*, a *right branch*, and a *central station*. In our examples, stations are underlined.

♠xxxxxx•♠•xxxx••♠••xx•••♠•••xx••♠••xxxx•♠•xxxxxx♠
{ Left Branch } { central } { Right Branch }

t-sphere for $d = 8$ and $t = 7$. Stations are underlined.

Figure 12: 'Stations', 'branches', and a 'central station'.

Definition 4.1 *Let $L_S = S - (x-1)d$. Let $R_S = S + (x-1)d$.*

So p is in L_S (R_S) iff p is the leftmost (rightmost) point of some t-sphere centered in S.

Lemma 4.1 *$p + x$ is either in L_S or in R_S.*
Proof: Consider the t-spheres centered in S. They form a disjoint cover of $\mathbb{Z}$. In particular, $p + x$ is an element of some t-sphere $S_t(q)$, $q \in S$. Suppose $p + x$ lies in the left branch of $S_t(q)$, but is not its leftmost element (see Fig.13). Then

$$p - d + x + \begin{cases} -2 & \in S_t(p) \\ -1, 0 & \notin S_t(p) \cup S_t(q) \\ 1 & \in S_t(q) \end{cases}$$

So $p_1 = p - d + x - 1$ and $p_2 = p - d + x$ are neither in $S_t(p)$ nor $S_t(q)$. Thus, p_1 and p_2 are covered by some other t-sphere(s) centered in S, and these t-sphere(s) do not intersect $S_t(p)$ and $S_t(q)$. How can that be? In a t-sphereall stations except the leftmost and the rightmost have length ≥ 3. Thus, p_1 and p_2 are the leftmost or the rightmost points of$t$-spheres centered in S. If p_1 or p_2 is the leftmost point of a t-sphere S' centered in S, then S' intersects $S_t(p)$ at $p + x - 1$, contradiction. So p_1 and p_2 are the rightmost elements of t-spheres $S_t(q_1), S_t(q_2)$ where $q_1, q_2 \in S$. Then $q_1 + 1 = q_2$, contradiction.

So if $p + x$ lies in the left branch of $S_t(q)$, $p + x$ must be its leftmost element. Else $p + x$ lies in the right branch of $S_t(q)$ or in its central station. Then by a similar proof $p + x$ must be the rightmost element of $S_t(q)$. □

```
♠xxxxxx•♠•xxxx••♠••xxxx•♠•xxxxxx♠
          ??♠xxxxxx•♠•xxxx••♠••xxxx•♠•xxxxxx♠
```

- $d = 8$, $t = 5$, $x = 3$.
- Upper row: $S_t(p)$; p and $p + x$ are encircled.
- Lower row: $S_t(q)$; $p - d + x - 1$, $p - d + x$ are labeled by '?'; q is encircled.

Figure 13: Proof of Lemma 4.1: $S_t(p)$ and $S_t(q)$.

Since the t-spheres centered in S are disjoint, $S_t(p + t)$ is the only t-sphere centered in S that contains $p + x$. Since $p + x$ is the inner point of $S_t(p + t)$, by Lemma 4.1 $p + t \notin S$. In particular, the two-offset construction from Section III could not be s-optimal since it started with {0,t,2t, ... , kt}.

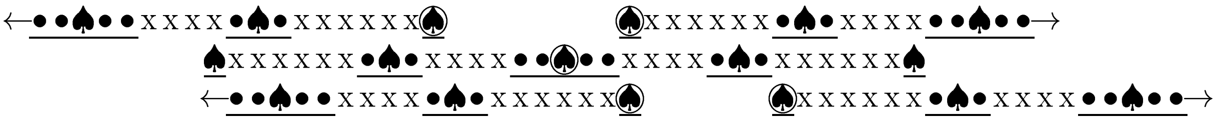
Lemma 4.2 *In each part, exactly one of the two statements is true (see Fig. 14):*

a) $p + x \in L_S$ and $p - d + x - 1 \in R_S$
   $p + x \in R_S$ and $p + d + x - 1 \in L_S$

b) $p - x \in L_S$ and $p - d - x + 1 \in R_S$
   $p - x \in R_S$ and $p + d - x + 1 \in L_S$

Proof: By Lemma 4.1, either $p+x \in L_S$ or $p+x \in R_S$. Suppose $p+x \in L_S$ and let $p' = p-d+x-1$. Since p' is neither in $S_t(p)$ nor in the t-sphere containing $p+x$, it is an element of some other t-sphere $T = S_t(q)$, $q \in S$. Since p' is the leftmost element of some station of T, $p+x = p'+d+1 \in T$, unless p' is the rightmost element of T.

Case $p+x \in R_S$ is solved similarly. Part (b) follows from part (a) by symmetry. $\square$



- $d=8$, $t=5$, $x=3$.
- 1st row: $p+x \in L_S$ and $p-d+x-1 \in R_S$ are labeled.
- 2nd row: $S_t(p)$ is shown; p is labeled .
- 3rd row: $p+x \in R_S$ and $p+d+x-1 \in L_S$ are labeled.

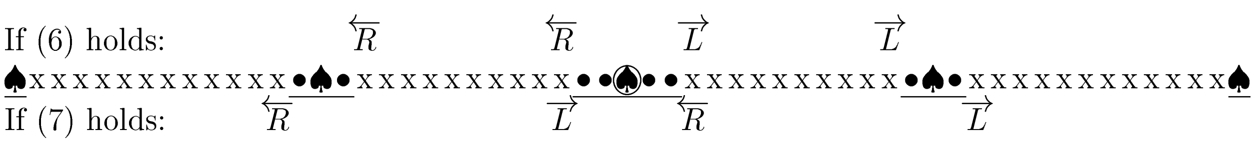
Figure 14: The two options in Lemma 4.2.

Lemma 4.3 *Exactly one of the following is true (see Fig. 15):*

$$p+x,\ p+d-x+1 \in L_S \quad \textit{and} \quad p-x,\ p-d+x-1 \in R_S \tag{6}$$

$$p+x,\ p-d-x+1 \in R_S \quad \textit{and} \quad p-x,\ p+d+x-1 \in L_S \tag{7}$$

Proof: By Lemma 4.2, there are four possible cases: (6), (7), $p \pm x \in L_S$, and $p \pm x \in R_S$. If $p \pm x \in L_S$ then by Lemma 4.2 $p_\pm \in S$, where $p_\pm = p - dx \pm (x-1)$. Thus $p_+ - p_- = t-1$, contradiction. Case $p \pm x \in R_S$ is ruled out similarly. $\square$



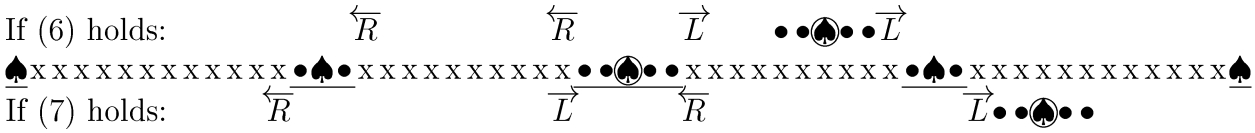
- The leftmost (L_S) and the rightmost (R_S) vertices are labeled by $\overrightarrow{L}$, $\overleftarrow{R}$, resp.
- p is encircled in the middle row.
- $d=14$, $t=5$, $x=3$.

Figure 15: The two cases in Lemma 4.3.

Lemma 4.4 *If (6) then $p+d-t \in S$. If (7) then $p+d+t \in S$ (see Fig. 16).*

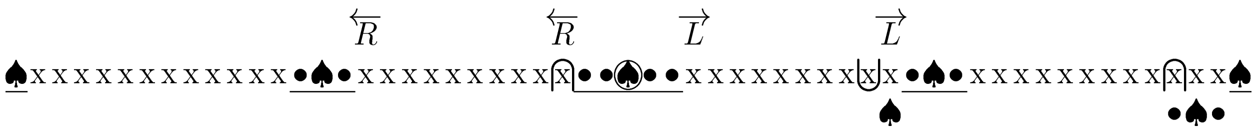
Proof: Suppose (6) holds. Let $T = S_t(q)$, $q \in S$, be the t-sphere containing $p' = p+d-x$ (Fig. 17a). Say p' is contained in the station W of T. Let W_L, W_R be the stations of T immediately to the left and immediately to the right from W. Unless W is the central station of T, either W_L or W_R is wider than W. Since p' is the rightmost point of W, this means that either $p-x \in W_L$ or $p+2d-x \in W_R$ (Fig. 17b). However, these two points belong to other t-spheres centered in S (Fig. 17c and 17d). Thus, W is the central station of T and $q = p + d - t$.

If (7), we let $T = S_t(q)$, $q \in S$, be the t-sphere containing $p' = p + d + x$. Then by a similar argument $q = p + d + t$. $\square$



- It is an elaboration of Fig. 15.
- In the upper row, $q = p + d - t$ is encircled; the central station of $S_t(q)$ is shown.
- Same for $q = p + d + t$ in the lower row.

Figure 16: The two cases in Lemma 4.4.



- $d = 14$, $t = 5$, $x = 3$.
- $S_t(p)$ is drawn in the middle row; p is encircled.
- Assuming (6) holds, the upper row shows the consequences of Lemma 4.3.

a) $p' = p + d - x$ is labeled by $\bigcup$.

b) $p - x$ and $p + 2d - x$ are labeled by $\bigcap$.

c) By Lemma 4.3, $p - x \in L_S$.

d) By Lemma 4.3, $p + d - x + 1 \in R_S$, i.e. it is the leftmost point of some t-sphere S' centered in S. We draw two leftmost stations of S' in the lower row to show that $p + 2d - x \in S'$.

Figure 17: For the proof of Lemma 4.4.

Theorem 4.1 *For odd t, s-optimal sets exist only if $d \equiv \pm t \pmod{|S_t|}$.*

Proof: Let S be an s-optimal t-sparse set. Take any $p \in S$. If (6) then by Lemma 4.4 $q = p + d - t \in S$. Now we apply Lemma 4.3 to q. Either (6) or (7) must hold for q. Since $q + x = p + d - x + 1 \in L_S$, (6) does. So we apply Lemma 4.4 again: $q + d - t \in S$. In the same fashion, $p + k(d - t) \in S$ for any $k \in \mathbb{N}$. Since this holds for any $p \in S$, S is periodic with a (not necessarily smallest) period $d - t$.

Let w be the number of points of S within one period. Since S is s-optimal, the density of S is $w/(d - t) = 1/|S_t|$. Thus, $|S_t|$ divides $d - t$.

If (7) holds for p, then by a similar argument $|S_t|$ divides $d + t$. $\square$

Theorem 4.2 *If t is odd and $d \equiv \pm t \pmod{|S_t|}$ then the even construction is s-optimal.*
Proof: Let $s = |S_t|$. Suppose $d \equiv t \pmod{s}$ (for $d \equiv -t$ the proof is similar). Suppose the even construction is not t-sparse. Then there are points $p > q$ s.t. $p \equiv q \pmod{s}$ and $dist(p,q) < t$. Let $\textsc{can}(p-q) = (i,j)$. Then s divides $id + j$, hence $it + j$. Since $it + j$ is smaller than $2s$, it is equal to s. Therefore, by a simple computation, $t = 2i \pm 1 = \pm(2j-1)$, so $dist(p,q) = i + |j| = t$, contradiction. $\square$

Case of even $t = 2x$

Recall definitions of *long* and *short* t-spheres from Section II. For odd t $dist(p,q) \geq t$ iff $S_t(p)$ is disjoint with $S_t(q)$. A similar lemma for even t is somewhat more complicated.

Lemma 4.5

$$\begin{aligned} dist(p,q) \geq t \quad &\iff \quad S_t^L(p) \textit{ is disjoint with } S_t^L(q) \textit{ and } |p-q| \neq t-1 \\ &\iff \quad S_t^S(p) \textit{ is disjoint with } S_t^S(q) \textit{ and } |p-q| \neq (t-1)d \\ &\iff \quad S_t^L(p) \cap S_t^L(q) = S_t^S(p) \cap S_t^S(q) = \emptyset \end{aligned}$$

As for odd t, we will consider the partitions of $\mathbb{Z}$ by t-spheres to investigate the s-optimal sets. We'll use *long* t-spheres. Again we find it convenient to talk about the branches of a t-sphere (see Fig. 18)

♠xxxxxx•♠•xxxx••♠••xxx••♠••xxxx•♠•xxxxxx♠
{ Left Branch } { Right Branch }

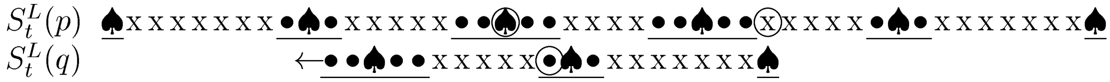
$d = 8$, $t = 6$. p and $p + d$ are labeled. Stations are underlined.

Figure 18: 'Stations' and 'branches' of $S_t^L(p)$.

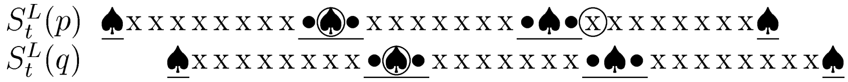
Lemma 4.6 *At least one of $p + x(d+1)$, $p - x(d-1)$ is in S.*
Proof: The long t-spheres centered S partition $\mathbb{Z}$. In particular, $p' = p + d + x$ is an element of some long t-sphere $T = S_t^L(q)$, $q \in S$. Clearly, p' is a leftmost element of some station of T. Which station? If p' is in the right branch of T then either $p + x - 1$ is in both T and $S_t^L(p)$ (Fig. 19a), or $q = p + t - 1$, which is too close to p (Fig. 19b).

So p' lies in the left branch of T. Now, if p' is the leftmost element of T then $q = p + x(d+1) \in S$, and we are done. Else $p + x - 1 \in S_t^L(p)$, $p + x + 1 \in T$, but $p + x$ is in neither (see Fig. 20). So $p + x$ must be either the leftmost or the rightmost element of some other long t-sphere $T' = S_t^L(q')$, $q' \in S$. It cannot be the leftmost element since in this case $p + d + x - 1$ is in both T' and $S_t^L(p)$. Thus, it is the rightmost element of T', in which case $q' = p - x(d-1)$. $\square$

Theorem 4.3 *For even t the even construction is s-optimal only if $d \equiv \pm 1 \pmod{t}$.*
Proof: The even construction is $S = \{kt^2/2 \mid k \in \mathbb{Z}\}$, where $t^2/2$ is the size of a t-sphere. So if S is s-optimal then by Lemma 4.6, $t^2/2$ divides either $(d+1)t/2$ or $(d-1)t/2$. $\square$

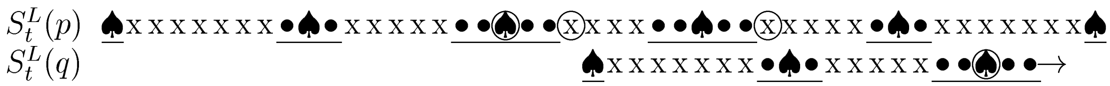


(a) $S_t^L(p)$ intersect T at $p+x-1$ (labeled in the lower row).
p and p' are labeled in the upper row. Here $d=9$, $t=6$.



(b) Labeled are: p, p' in the upper row, q in the lower row.
p and q are too close. Here $d=10$, $t=4$.

Figure 19: For the proof of Lemma 4.6: what if p' lies in the right branch of T?



Labeled are: p, $p+x$, p' in the upper row, q in the lower row.
What do you cover $p+x$ with? Here $d=9$, $t=4$.

Figure 20: For the proof of Lemma 4.6: p' must lie in the left branch of T

Discussion and open questions

For odd t we know exactly when s-optimal constructions exist. For even t we just have Theorem 4.3 about s-optimal *even* construction. Since for any $d \equiv \pm 1 \pmod t$, t even, there exists an s-optimal two-offset construction, we conjecture that Theorem 4.3 actually holds for *all* t-sparse sets.

Now back to the even construction for even t. When exactly is it s-optimal? At this point it is an open question; unfortunately, the converse to Theorem 4.3 is false. How much do we care, really? By Theorem 4.3 if the even construction is s-optimal, there exists an s-optimal two-offset construction. However, the even construction is "nicer", so we'd like to use it whenever we can. For this reason we investigated this question further.

Let D_t be the set of all values of d s.t. the even construction is s-optimal. We computed $\min(D_t)$ and the first 20-30 elements of D_t for each $t \leq 42$. This data motivated several conjectures:

- Let p be the smallest prime that does not divide $t/2$. Then $\min(D_t) = pt-1$.
- Let $d \equiv 1 \pmod t$. Then $d \in D_t$ iff $d-2 \in D_t$.
- Consider the sequence of intervals between consecutive elements of D_t. This sequence is periodic, starting from the very first element of the sequence. Let $p_0, p_1, \ldots, p_n$ be the distinct prime divisors of $t/2$. Then the length of the period is $2 \times \prod_{j=0}^{n}(p_j - 1)$, and the sum of the elements in a period is $t \times \prod_{j=0}^{n} p_j$.

5 Case $t > \lceil d/2 \rceil$: Constructions and Lower Bounds

In this section we consider the case $t > \lceil d/2 \rceil$. For $t > d-2$ there is a simple unique optimal interleaving scheme and a t-sparse set. For each choice of (d,t) s.t. $\lceil d/2 \rceil < t \leq d-2$ we present a family of optimal t-sparse sets and extend one of them to an interleaving scheme that is optimal in about half of the cases and a $(1+\frac{4}{t})$-approximation otherwise.

Recall the definitions of $v^{\tau}_{\max}$ and $v^{\tau}_{\min}$ from section II. Let $v_{\max} = v^{t}_{\max}$, $v_{\min} = v^{t}_{\min}$. Let $\delta = \lceil d/2 \rceil$, $\tau = \lceil t/2 \rceil$.

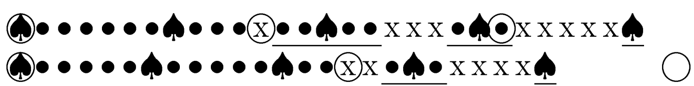
Definition 5.1 *Let the even labeling $\mathcal{E}$ be a labeling on $\mathbb{Z}$ with a period $\{1, 2, 3, \ \ldots \ , v_{min}\}$.*

Theorem 5.1 *$\mathcal{E}$ is an interleaving scheme iff $t > d-2$, in which case it is the unique optimal interleaving scheme.*

Proof: By Lemma 2.6, if $t \leq d-2$ then $2v_{\min}$ is t-reachable, so $\mathcal{E}$ is not an interleaving scheme (Fig 5a). Now suppose $t > d-2$. By Lemmas 2.5 and 2.6, $v_{\max} < 2v_{\min}$. Since the distance between any two points labeled the same in $\mathcal{E}$ is either $v_{\min}$ or greater than $v_{\max}$, $\mathcal{E}$ is a valid interleaving scheme (Fig 5b).

To prove that $\mathcal{E}$ is unique optimal, we show that any other interleaving scheme requires more labels. Indeed, in any other interleaving scheme there are two consecutive vertices u,v labeled the same s.t. $|u-v| > v_{\min}$. Let S be the integer interval $[u+1; u+v_{\min}]$. By definition of $v_{\min}$, the distance between any two points in S is less than t. Therefore each point in S must be assigned a distinct label different from the label of u and v, which requires at least $v_{\min}+1$ labels. □

a)
b)



(a) $t \leq d-2$; (b) $t > d-2$. Vertices 0, $v_{\min}$, $2v_{\min}$ are labeled.

Figure 21: $\mathcal{E}$ is an interleaving scheme iff $t > d-2$.

Corollary 5.1 *$v_{min}\mathbb{Z}$ is t-sparse iff $t > d-2$, in which case it is the unique optimal periodic t-sparse set.*

Proof: The second assertion holds because the interval between any consecutive elements of a t-sparse set is at least $v_{\min}$. □

Note that the set $v_{\min}\mathbb{Z}$ is exactly the greedy construction from Section III applied to the case $t > d-2$.

For the rest of this section, assume $\delta < t \leq d-2$. Equivalently,

$$v_{\min} \ = \ (t-\delta)d + \delta \ \leq \ v_{\max}/2$$

Definition 5.2 *Call a point w remote if $dist(w) \geq t$. Call (w_1, w_2) a remote pair if w_1, w_2 and $w_1 + w_2$ are positive and remote. Call $w_1 + w_2$ the sum of a remote pair. Call a remote pair minimal if its sum is minimal for given (d,t). Define a* remote-pair set *$\mathcal{R}^{w_2}_{w_1}$ as a periodic set S with a period $w_1 + w_2$ s.t. $S \cap [0, w_1 + w_2) = \{0, w_1\}$.*

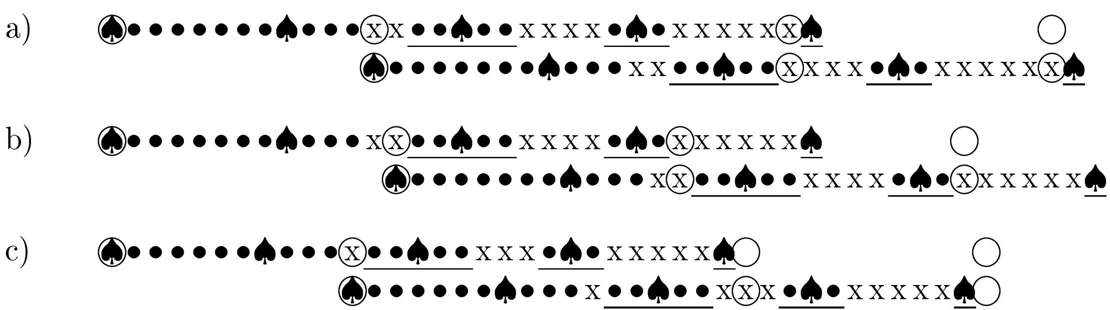
We will derive a *Remote-Pair* Lower Bound (RLB) on $\text{IND}_{\min}$ from a lower bound on the sum of a remote pair. If a set is t-sparse and its index reaches RLB, call it *r-optimal.* We will prove that for any minimal remote pair (w_1, w_2) the set $\mathcal{R}_{w_1}^{w_2}$ is r-optimal

Lemma 5.1 *Fix (d,t). If the sum of any remote pair is at least w, then* $\text{IND}_{min} \geq w/2$.
Proof: Let S be a t-sparse set with a period p. Let $s_0 \ldots s_{2n}$ be the elements of $S \cap [0, 2p]$ in increasing order. Then for each i $(s_{2i+1} - s_{2i},\ s_{2i+2} - s_{2i+1})$ is a remote pair, so its sum $s_{2i+2} - s_{2i}$ is at least w. Therefore $s_{2n} - s_0 \geq wn$, so

$$\text{IND}(S) = \frac{s_{2n} - s_0}{2n} \geq w/2$$

□

Let $v_0 = v_{\min}$. For even d, define $v_1 = v_{\min} + 1$. Let σ_i be the minimal sum of a remote pair $(v_i,\ \cdot)$ (Fig 22). Define $\sigma_{\min}$ as σ_0 for odd d, and as $\min(\sigma_0, \sigma_1)$ for even d. Note that v_1 and σ_1 are not defined for odd d.



- In each example, for $i = 0$ or 1,
  - t-spans are depicted, in the notation explained in Section II.
  - the upper line is a t-span of 0.
  - the lower line is a t-span of σ_i.
  - σ_i is computed as the leftmost point that is remote in both lines.
  - points 0, v_i, σ_i and $\sigma_i + v_i$ are labeled.
- In example (a), $(d,t) = (8,5)$; σ_0 is computed.
- In example (b), $(d,t) = (8,5)$; σ_1 is computed.
- In example (c), $(d,t) = (7,5)$; σ_0 is computed.

Figure 22: Computing σ_0 and σ_1.

Theorem 5.2 (Remote-Pair Lower Bound) $\text{IND}_{\min} \geq \sigma_{\min}$.
Proof: By Lemma 5.1 it suffices to prove that the sum of any remote pair is at least $\sigma_{\min}$. Consider the following loop.

```
FUNCTION FOO
INPUT: remote pair (w1,w2);
LET w1≤w2, CAN(w1)=(x,y);
WHILE |y|<δ or dist(w1)>t
```

```
      IF dist(w)>t LET z=d
          ELSE IF y<0 LET z=d+1
          ELSE LET z=d-1;
      (w1,w2)←(w1-z, w2+z);
  RETURN (w1,w2).
```

By Lemma 2.4 after each iteration w_1, w_2 remain remote, so, since $w_1 + w_2$ is invariant, (w_1, w_2) remains a remote pair. The loop ends because if $dist(w_1) > t$ then $dist(w_1)$ decreases, else $|y|$ increases.

Let (w_1, w_2) be a remote pair. Let $(w_1', w_2') = \mathtt{FOO}(w_1, w_2)$. Then (w_1', w_2') is a remote pair s.t. $dist(w_1') = t$ and $|y| = \delta$, where $\textsc{can}(w_1') = (x, y)$. Therefore, w_1' is v_0 or v_1. By definition of σ_0 and σ_1, $w_1 + w_2 = w_1' + w_2' \geq \sigma_{\min}$. □

Definition 5.3 *Call a remote pair* standard *if its sum is σ_0 or σ_1.*

We will see that for each pair (d, t) there exists a family of standard remote pairs of the form $(v_i + j(d+1),\ \sigma_i - j(d+1))$.

Definition 5.4 *Call $\mathcal{R}_{w_1}^{w_2}$ an* s-remote-pair set *if (w_1, w_2) is a standard remote pair.*

Theorem 5.3 *Any s-remote-pair set is t-sparse.*
Proof: Let (w_1, w_2) be a standard remote pair. For any $u, v \in \mathcal{R}_{w_1}^{w_2}$, either

$$|u - v| \in \{0, w_1, w_2, w_1 + w_2, 2w_1 + w_2, w_1 + 2w_2\}$$

or else $|u - v| \geq 2(w_1 + w_2) \geq 2\sigma_{\min}$, so $dist(v, u) \geq t$ by Lemma 5.2. Therefore, it remains to prove that $2w_1 + w_2$ and $w_1 + 2w_2$ are remote, which follows from Lemma 5.3. □

Lemma 5.2 $2\sigma_{min} > v_{max}$.

Lemma 5.3 *$\sigma_0 + v$ and $\sigma_1 + v$ are remote for any remote $v > 0$.*

By Thm 5.3, $\mathcal{R}_{v_0}^{\sigma_0 - v_0}$ is exactly the greedy construction from Section III applied to the case $\delta < t \leq d - 2$. The following chain of lemmas proves Lemmas 5.2 and 5.3 which are required to complete the proof of Thm 5.3. To simplify formulas, define γ as 0 if d is odd and 1 if d is even. For $i \in \{0, 1\}$ define

$$\begin{aligned}
(\alpha_i,\ \beta_i) &= \textsc{can}(\sigma_i - v_{\min} - v_i) \\
(\alpha,\ \beta) &= \textsc{can}(v_{\min} + v_i) \\
&= (2(t - \delta) + 1,\ i + 1 - \gamma)
\end{aligned}$$

Note that $\beta \in \{0, 1\}$. We are slightly abusing the notation since β depends on i.

Fact 5.1 *For a positive $v < v_{max}$ the following are equivalent:*

a) v is remote.

b) $\textsc{can}(v - v_{min}) = (\mu_1, \mu_2)$, where $-\mu_1 \leq \mu_2 \leq \mu_1 + \gamma$.

c) For some $i \in \{0, 1\}$ $v = v_i + \mu_1 d + \mu_2$, where $|\mu_2| \leq \mu_1$.

Corollary 5.2 *Since $\sigma_i - v_i$ is remote, $-\alpha_i \leq \beta_i \leq \alpha_i + \gamma$.*

Lemma 5.4 $\textsc{can}(\sigma_i) = (\alpha + \alpha_i,\ \beta + \beta_i)$.

Proof: Since $\sigma_i = (\alpha + \alpha_i)d + (\beta + \beta_i)$ and $\beta \in \{0, 1\}$, the Lemma holds unless $\beta = 1$ and $\beta_i = \delta$. In the latter case, let $\sigma = \sigma_i - 1$. Then σ is remote (since $\sigma \geq v_{\min}$ and $\sigma \equiv \delta \pmod{d}$), and $\sigma - v_i$ is remote by Fact 5.1b. So (v_i, σ) is a remote pair, which contradicts the minimality of σ_i. □

Lemma 5.5 $\alpha_i = \delta - \lceil (t+i)/2 \rceil$. $\beta_i = \alpha_i + \gamma$.

Proof: Unfortunately, we need more notation:

$$\begin{aligned} (\alpha_i^\sigma,\ \beta_i^\sigma) &= \textsc{can}(\sigma - v_{\min} - v_i) \\ W_i &= \{\sigma > v_i |\ -\alpha_i^\sigma \leq \beta_i^\sigma \leq \alpha_i^\sigma + \gamma \text{ and } -\delta < \beta + \beta_i^\sigma \leq \delta\} \end{aligned}$$

Then by Fact 5.1b and Lemma 5.4 σ_i is the smallest remote element of W_i. Let $\varphi(x)$ be the maximal distance between 0 and $\sigma \in W_i$ s.t. $\alpha_i^\sigma = x$. Since for each $\sigma \in W_i$

$$dist(\sigma) = (\alpha + \alpha_i^\sigma) + |\beta + \beta_i^\sigma|,$$

$\varphi(x) = 2x + \alpha + \beta + \gamma$ (which requires $\beta_i^\sigma = \alpha_i^\sigma + \gamma$), so the Lemma follows since $\alpha_i = \min\{x |\ \varphi(x) \geq t\}$. □

Corollary 5.3 $\sigma_i - v_i = v_{min} + \alpha_i(d+1) + \delta$.

Proof of Lemma 5.2: By Lemma 5.5, $\alpha + \alpha_i \geq \tau$. So $\sigma_i > \tau d - \delta + 1$. Thus, $2\sigma_i > (t-1)d = v_{\max}$. □

Lemma 5.6 *Suppose u, v are positive and remote, and $\textsc{can}(u) = (x, y)$ s.t. $|y| \leq t/2$. Then $u + v$ is remote.*

Proof: Wlog assume $v < v_{\max}$. Since $dist(u) = x + y \geq t$, $x \geq t/2$. By Fact 5.1c

$$u + v = v_i + (x + \mu_1)d + (y + \mu_2)$$

where $i \in \{0, 1\}$ and $|\mu_2| \leq \mu_1$. Since $|y + \mu_2| \leq x + \mu_1$, by Fact 5.1c $u + v$ is remote. □

Proof of Lemma 5.3: Use Lemma 5.6 since by Lemma 5.5 $|\beta + \beta_i| \leq t/2$. □

This concludes the proof of Thm. 5.3. Now we will extend s-remote-pair sets to interleaving schemes. An efficient covering of $\mathbb{Z}$ by copies of a remote-pair set is given by the following analogue of Thm. 3.2.

Lemma 5.7 *Let $S = \mathcal{R}_{w_1}^{w_2}$ be a t-sparse remote-pair set s.t. $w_1 < w_2$. Let $g = \gcd(w_1, w_2)$. Then*

$$\textsc{deg}(S) = \begin{cases} \textsc{ind}(S), & g = 1 \\ \textsc{ind}(S) + g, & \textit{otherwise} \end{cases}$$

Proof: Use same arguments as in Thm. 3.2. □

Corollary 5.4 *Standard remote pairs that are relatively prime yield r-optimal interleaving schemes.*

Reasoning about gcd's of *all* standard remote pairs is quite a nightmare. Instead, for each choice of (d,t) we will come up with a single s-remote-pair set that yields a good interleaving scheme. Let

$$\alpha^* = \begin{cases} \delta-\tau, & d \text{ or } t \text{ is odd} \\ \delta-\tau-1, & d \text{ and } t \text{ are even} \end{cases}$$

$$(w_1^*, w_2^*) = \begin{cases} (v_0, \sigma_0 - v_0), & \text{odd } d \\ (v_1, \sigma_1 - v_1), & \text{even } d \end{cases}$$

Then by Thm. 5.5 (w_1^*, w_2^*) is a minimal remote pair, with a sum $\sigma_{\min}$. By Cor. 5.3 $w_2^* - w_1^* = \alpha^*(d+1)$. Therefore for each $j \leq \alpha^*/2$,

$$(w_1^* + j(d+1),\ w_2^* - j(d+1))$$

is a minimal remote pair. In particular, there is a minimal remote pair of the form (w^*, w^*) if α^* is even, and $(w^*, w^* + d + 1)$ if α^* is odd.

Definition 5.5 *Define the* m-remote-pair set $\mathcal{R}$ *by*

$$\mathcal{R} = \begin{cases} w^*\mathbb{Z}, & w^* = \frac{\sigma_{min}}{2}, \quad \text{if } \alpha^* \text{ is even,} \\ \mathcal{R}_{w^*}^{w^*+d+1}, & w^* = \frac{\sigma_{min}-d-1}{2}, \quad \text{if } \alpha^* \text{ is odd.} \end{cases}$$

By the previous discussion $\mathcal{R}$ is r-optimal. If α^* is even, $\mathcal{R}$ obviously extends to an r-optimal interleaving scheme. If α^* is odd, with some arithmetic one shows that $\gcd(w^*, d+1) = \gcd(d+1, t)$, so by Lemma 5.7 $\mathcal{R}$ extends to an interleaving scheme that is r-optimal if $\gcd(d+1,t) = 1$, and a $(1+\frac{4}{t})$-approximation of RLB otherwise.

Lower bounds on non-periodic constructions

RLB can be extended to non-periodic interleaving schemes and t-sparse sets by the following version of Lemma 5.1.

Definition 5.6 *Define the density $\rho(S)$ of a set S as a limit*

$$\rho(S) = \lim_{n\to\infty} \frac{|S \cap [-n; n)|}{2n}$$

If this limit exists, call S regular and define the index IND(S) *of S as the inverse of density rounded up.*

Lemma 5.8 *Fix (d,t). If the sum of any remote pair is at least w then* DEG $\geq w/2$ *for any interleaving scheme, and* IND $\geq w/2$ *for any regular t-sparse set.*

Proof: Let w_1, w_2, w_3 be three consecutive vertices labeled the same in an interleaving scheme. Then $(w_2 - w_1, w_3 - w_2)$ is a remote pair, so its sum $w_3 - w_1$ is at least w. Therefore, in the integer interval $[0; w)$ at most two vertices can be marked by each label, which requires at least $w/2$ distinct labels.

Let S be an infinite regular t-sparse set. Let $\{s_i\}$ be the an increasing enumeration of S. For each i, $(s_{i+1} - s_i,\ s_{i+2} - s_{i+1})$ is a remote pair, so its sum $s_{i+2} - s_i$ is at least w. Then $s_n - s_{-n} \geq nw$ for any $n > 0$. This gives $\text{IND}(S) \geq w/2$ since

$$\rho(S) = \lim_{n\to\infty} \frac{2n}{s_n - s_{-n}} \leq 2/w$$

□

Similarly, Cor. 5.1 can be strengthened:

Lemma 5.9 *If $t > d - 2$ then $v_{min}\mathbb{Z}$ is the unique optimal* regular *t-sparse set.*

References

[1] M. Blaum, J. Bruck and A. Vardy, "Interleaving schemes for multidimensional cluster errors," *IEEE Transactions on Information Theory*, vol.44, No.2, pp. 730-743, March 1998.

[2] M. Blaum, J. Bruck and P.G. Farell, "Two-dimensional interleaving schemes with repetitions," *IBM Research Report RJ 10047*, Oct. 1996.

[3] B. Elspas and J. Turner, "Graphs with circulant adjacency matrices," *J. Combinatorial Theory*, No.9, pp. 297-307, 1970.

[4] T. Etzion and A. Vardy, "Two-dimensional interleaving schemes with repetitions: constructions and bounds," *IEEE International Symposium on Information Theory*, 2000.